

⑩ 日本国特許庁(JP)

⑪ 特許出願公開

⑫ 公開特許公報(A)

昭60-123000

⑬ Int. Cl.<sup>4</sup>

識別記号

庁内整理番号

⑭ 公開 昭和60年(1985)7月1日

G 10 L 5/06

GLA

8221-5D

審査請求 未請求 発明の数 1 (全9頁)

⑯ 発明の名称 話者に影響を受けない話者認識方法

⑰ 特 願 昭59-234856

⑱ 出 願 昭59(1984)11月7日

優先権主張 ⑲ 1983年11月8日 ⑳ フランス(FR)㉑ 8317738

⑳ 発 明 者 ジェラルド ビクトー フランス国サン ポール、ルート ドウ・ラ コル、1185  
ルベンバサ㉒ 出 願 人 テキサス インスツル アメリカ合衆国テキサス州ダラス、ノース セントラル  
メンツ インコーポレ エクスプレスウェイ 13500  
イテッド

㉓ 代 理 人 弁理士 浅 村 皓 外2名

## 明 細 書

## 1. 発明の名称

話者に影響を受けない話者認識方法

## 2. 特許請求の範囲

(1) 分析する音声を音声ユニットに分解し

辞書手段の中に含まれる音素とこれらの音声ユニットとを比較する工程を含む不特定話者の音声認識方法において、試験期間の間に音素とこれに対応する領域との間の関連可能性を計算することによつて上記音声空間の各々の領域に音声ユニット表示を付与することから成る音声ユニットの検証工程を含むことを特徴とする上記音声認識方法。

(2) 上記方法において、任意の点を音声空間の領域と関連づけ、さらに所定数の音声ユニットと関連づけるために使用される最大の距離が充分広い領域の数を選択することによつて必要なだけ短縮されることを特徴とした特許請求の範囲第1項に記載される方法。

(3) 上記方法に於て音声信号は、ベクトル量子化装置(1)で分析され、上記装置では、所定の値

の時間間隔ごとに音響パラメータが計算され距離を計算することによつて辞書(2)の中に含まれる各々のテンプレートと比較されることを特徴とする特許請求の範囲第1項又は第2項に記載される方法。

(4) 上記距離の計算はユークリッド距離の計算であることを特徴とする特許請求の範囲第3項の方法。

(5) ベクトル量子化装置(1)の出力から得た列が各々のテンプレートに関連する音声周波数のテーブル(3120)を用いて音素形式で記憶される単語の辞書(5)と比較されることを特徴とする特許請求の範囲第3項又は第4項に記載される方法。

(6) ダイナミックプログラミングアルゴリズムが上記ベクトル量子化装置の出力から得た列と辞書内の各々の単語との整合可能性を計算する為に使用されることを特徴とする特許請求の範囲第5項の方法。

(7) 上記整合可能性の計算は、音声周波数テーブ

ルの使用に基づいて行われることを特徴とする特許請求の範囲第6項の方法。

(8) 上記方法がスペクトルテンプレート(9)の辞書内で最も近い近隣のスペクトルテンプレートからの距離が閾値より大きくなるようなスペクトルテンプレート(10)を含む試験用語句の組(8)を選択し、上記試験用語句の組のスペクトルテンプレート(10)は、辞書内で最も近い近隣のテンプレートとともにクラス(11)にクラスター化され、辞書(9)内に登録するため各々のクラスの重心を選択する工程を含み、この工程が平均距離が所定閾値より小さくなるまで又は、変化が非常にわずかな値より小さくなるまで反復されることを特徴とする特許請求の範囲第1項から第7項までのいずれか1つの方法。

(9) 登録のための試験用工程が異なるアクセント及び音調を持つ所定数の話者によつて発声され、デジタル信号に変換された試験用語句の組の音声から分析結果ファイル(25)と音楽表示ファイル(29)を形成し、上記分析結果ファイル及び音

音楽表示ファイルから周波数テーブル(33)を形成する特許請求の範囲第1項から第8項のうちのいずれかに記載される方法。

(10) 上記分析結果ファイルは、試験用語句の組の音声信号を変換することによつて得たデジタル信号を符号化(22)し、上記デジタル信号を離形予測法による分析方法で分析することによつて形成され、一方、分析結果ファイル(25)のデータを音楽表示付与オペレーションにかけることによつて音楽表示ファイル(29)を形成することを特徴とする特許請求の範囲第1項から第9項までのいずれか1つに記載される方法。

(11) 上記音楽表示付与オペレーションは、上記試験用語句の組から得たデジタル音声信号を示すための可聴器(27)及び同時にかつ一時的に上記信号を表示するグラフ表示手段(28)とを使つて音楽の一定している部分のフレームを決定することを特徴とする特許請求の範囲第9項の方法。

(12) 辞書インデックス(31)内に最適なテンプレートを配置することによつて検索インデックス

(25)及び分析インデックス(29)から最適なスペクトルテンプレートを選択し辞書インデックス(31)の内容から周波数を計算することによつて周波数テーブルを作ることとを特徴とする特許請求の範囲第8項から第10項のいずれか1つの方法。

### 3. 発明の詳細な説明

#### <産業上の利用分野>

本発明は、音声認識システムに関し、さらに詳しくは、不特定話者の音声を検出する装置に関する。

#### <従来の技術及び発明が解決しようとする問題点>

音声認識の分野では、話者に対する独立性の問題が解決すべき重要な問題である。

特定話者の音声認識には、非常に有効である今日一般に知られる音声認識プロセスは、不特定話者の音声認識に用いる場合には、必要となる情報処理システムの負担が大きく、満足のゆく解決方法を提供することはできない。

実際は、話者の特性に左右されない音声認識シ

ステムを提供する為、認識すべき単語に対し10個のテンプレートデータを必要とする方法が既に発表されている。(これは、1979年4月発行のIBMトランザクション オブ A B D P 27巻2号掲載のT. B. レナンソン他「個々の単語認識のための不特定話者に対し参照テンプレートを選択する音声相互結合技術」に示されている。)

この技術では、データ処理上の負担及びランダムアクセスメモリ(RAM)のサイズが既にかなり増大することになり、さらに、困ったことには、このテンプレートデータは、多数の注意深く選択した話者によつて発声される多数の語句から抽出して得たものでなくてはならなかつた。

この為に要求にあわせて語いを構成することと非常に大変な、登録のための作業が必要となつた。これによつてこのようなシステムを使う可能性は、限られてしまう。

更に、音声入力を音声ユニット列に変換しようとする分析による解決策で、いまだに高いレベルの成果を得たものはない。

この理由としては以下の点が考えられる。

いくつもの開発計画において、分析による手法が連続する音声の認識に関し、利用されてきた。(1977年5月のプロシーディング1977 IEEE インターナショナル コンファレンス オン ASSP ハートフォードシタイ、799頁から802頁に掲載される L. D. エルマンの論文「HEARSAY システムの機能的説明」及び1978年11月13～15日2巻コンgres オブ ザ ACOUSTIC フット シフター イベントの304～314頁に掲載されるメリルその他の論文「音声対話のためのシステム」に示される。)

当然ながら、これによつてその他の困難な問題点が加わる。

重大な問題点としては、使用される語いに制限のない(又は、語いが非常に広い)場合には音声ユニットへのセグメンテーションを必要とし、この段階で起こるエラーをとり除くことは非常に困難であるという点があげられる。

<問題点を解決するための手段及び作用>

これによつて使用される音響距離が各音の間に知覚される距離と密接に関連することが仮定される。

最短距離を見つけることは、あまり難しくはないが、最短でなく長めの距離に関しても関連性を認めることができる関連を示す距離を持つ音声空間には残念ながら知られていない。

故に従来の方法を使用すると、音声距離が常に充分に短くなるように、より適当な音素表示又は満足のゆくテンプレートをもつてみつけださなくてはならない。

これらは、どちらにしても、同じ位、困難である。

このような困難な作業を省くための簡単な方法としては、本発明に従うと、音声空間のいずれの領域も数個の音声ユニットを表示するようにすることが提案される。

さらに、各々の領域において音声ユニットと関連する可能性が試験用語句を発声した音声から計算される。

従つてこの問題を解決するため、本発明では、特に語いが限られる分析方法を使つてダイナミック プログラム アルゴリズムによつて語いの中の各単語に対し最適なセグメンテーションを行うことを提案するものである。

他の問題点は、見つけるべき音声ユニットを示す音声テンプレートを作りだすことに関して生じる。

通常は、上記テンプレートは、任意の音声空間(acoustic space)に音声ユニットをマッピングし、各々の音声空間について最適な代表領域を見つけることによつて得られる。さらにこれらの領域の重心は、それぞれ対応する音声ユニットと結びつけられる。

(1人の話者についてのみならず多数の異なる話者についてさえも)各領域の多くは重なりありなので、この音声空間に通常一点が決まり、これによつて音声ユニットを表示している可能性が計算される。この可能性は、対応する領域の重心からこの音声ユニットまでの距離と比例する。

領域は、音響基底(acoustic basis)によつて、任意の数の領域を設定することによつて決定される。

故にこれらの領域によつて望み通りの正確さで「枠」を音声空間内に設定する。

任意の点を与えられた領域と結びつけ従つて所定数の音声ユニットと結びつける為に使用される最大距離は、必要を限り短くすることができる。

この技術を使うことによつて音声距離を厳密に認識と関連させる必要性は従来よりずっと少なくなり各々の音声ユニットに対し多数の正確なテンプレートを設定する必要はなくなつた。

この様な解決方法を実行するには、以前より多くの領域を使うことになる。しかしながら、これによつて、不特定話者の音声認識装置の能力がかなり向上される。

<実施例>

第1図の装置は、ベクトル量子化装置1を有する。ベクトル装置は、入力の中の1つで分析が行われる音声信号を受けとっている。

ベクトル量子化装置1の他の入力は、メモリ2に接続している。メモリ2は、スペクトルテンプレートの辞書データを記憶する。各々のテンプレートと関連する音楽の分布テンプレートのグループを記憶する装置の各グループは、各々のテンプレートと関連を持つ。

量子化装置1の出力及び音声分布テンプレートのグループ記憶装置3の出力は、ダイナミックプログラミングによつて、その単語である可能性を決定する装置4のそれぞれ対応する入力に接続される。この装置4は、音楽の形式で単語辞書を記憶しているメモリ5と共働する。

単語を発声した音声信号は、ベクトル量子化装置1に於て分析される。ベクトル量子化装置1では、音響パラメータ(この中では、線形予測法によつて得られるスペクトル係数)が、10~20msの時間フレームごとに計算される。このパラメータは、この場合ではコータリッド距離を用いて距離の計算が行われ、辞書の各々のテンプレートについて比較し、最も近いものを見つげだされ

る。故に、音声入力信号はテンプレートアドレスの列に変換される。

この辞書のテンプレート列は、単語辞書5に記憶される各々の単語と比較され、単純な音声表示の形式で記憶される。

音声表示と列内に含まれる辞書のテンプレートとの関連する可能性を用いることでダイナミックプログラミング アルゴリズムによつて、特定されるべき上記テンプレート列を辞書5内の音楽データに対応させる関連可能性の計算が可能になる。

関連可能性が所定の閾値以上の値であれば、最も高い可能性を持つ単語が話者によつて発音された単語であるとして選択される。

ダイナミックプログラミング アルゴリズムによつて単語辞書5内の各々の単語とマッチさせる最適な積分路を見つけている。このマッチング工程は、単語マッチング装置4に於て行われる。

この工程は、第2図のグラフで示される。この図で横軸は特定すべき辞書内の単語縦軸は、音楽

を示す。

最適積分路は、おれ離れで示される。ここではDPマッチングを行う上で考慮しなくてはならない最適パスを探すための領域制限の枠が7で示される。

この枠を使い技術によつて与えられた1つの音楽片に関連するテンプレートの数を制限しなくてもよくなる点に注意してほしい。しかし音楽の一定している部分も変化している部分もこの枠をとることができる。

マッチングの可能性が最も高いパス8は、ダイナミックプログラミングによつて決定される。

比較工程ごとに(エ軸上に)テンプレートによつて示される音楽片が(γ軸上の)音楽の一部である可能性が関連する表1の中から見つげられる。

7で示す点に到達するためには、DPパス制限の枠7によつて最適パス8は点7aからすぐ左にのびるか(従つて水平にのびる部分のパスとなる)又は対角線上の点7bからのびるか(従つて垂直にのびるパス部分を作ることがないよう)

のいずれかに決まる。

これらの各々のパスのマッチング可能性は、点7aからのパスと点7bからのパスの各々の可能性をテーブルTでみつけた点7の可能性を掛けることによつて計算される。本実施例では、音楽片7eが音楽0を示す可能性はP1である。

7に到達する路として選択される路は、最大の可能性値を持つパスである。

行列内に存在しうる各点に対しこのような工程を行つてゆくと最大の可能性値を持つパス8になり。

水平方向にのびるパスの長さは、DPマッチング制限の枠によつて制限されることはないが、実際には、この長さは、可能性の計算をする上で列内の音楽片と与えられた音楽との関連が上記関連可能性と無関係でないことも考え合わせて各々の音楽の長さを統計につた分布を用いることで制限される。

そこで第3図に関連して辞書手段の構造とその組合せのための試験工程を示す。

音響テンプレートの辞書は、同じ数の音楽を含み、多数の異なる話者によつて発音された試験用の句のグループをもとに線形予測符号化法によつて分析を行い作りだされる。

辞書が、平均的ゆがみを計測し、これに基づき、与えられた数のテンプレートに関し、最もよく試験用語句のグループを示す表示を提供できるようにテンプレートが選択される。

第3図は、ダイナミックグループ アルゴリズムに従う選択方法を示す。第3図では、試験用語句のグループ8とスペクトルテンプレート辞書9が図示される。選択オペレーションは、試験用語句のグループ8の中のスペクトルテンプレート10を選択することから始まる。ここでは、辞書内で最も近い近隣のテンプレートからの距離が所定の閾値以上であるスペクトルテンプレート10が選択される。辞書のサイズは、ただ辞書を作る為のプロセスを適当に中断することによつて適み通りに大きさを制限できる。

故に、試験用語句のグループに含まれる全ての

スペクトルテンプレートは、辞書内で最も近い点を使つてクラス11内にクラスター化される。その後、各々のクラス11の重心(クラスターセンター)がテンプレートとして選択される。

同様の工程が平均距離が所定の閾値より小さくなるまで、または距離の変化が計算処理が収束したことを示す非常に小さい値に達するまでくり返し続けられる。

故に、スペクトルテンプレート13を含む辞書9は、試験用語句のグループ内のスペクトルテンプレートのクラス11の重心12の値によつて構成される。

第4図の行14は、文を音楽で示した表示の列を示す。

行15は、変化する部分と変化しない部分に分けるマニュアルによるセグメンテーションを示す。行16は線形予測符号化音楽片を示す。

これらの符号化音楽片17は、スペクトルテンプレート辞書18内にこれに対応する音楽片13を有している。これらの音楽片のマッピングは、

スペクトル距離を用いてベクトルを符号化することによつて行われる。

最後に、第4図の下の方には、各々のスペクトルテンプレートに関連する音楽分布テーブル20が示されている。これらのテーブルが構成される段階を示す第2図のテーブル21にこれらのテーブルが相当する。

(第3図に示す)各々のクラス11に現われる音楽の数を数えることによつてこれらの音声周波数テーブルが得られる。

このことは試験用語句のグループが音楽にセグメンテーションされ表示がつけられたことを意味する。第1の試験用語句のグループに関しては、上記オペレーションは全くマニュアル操作で行わなくてはならないが、「満足ゆく程度」の辞書が作られた後は、認識アルゴリズム自体を使つて、他の試験用語句の自動的なセグメンテーションが可能となりマニュアル操作は、その検証のみになる。

辞書形成の為の試験工程のオペレーションは、

第5A図及び第5B図を参照して説明する。

第5A図は音楽表示をつけるオペレーションを図示する。

これらのオペレーションは、試験用語句のグループの音声のアナログからデジタルへの変換21及び線形予測法によるデジタル信号の分析23工程を含む。デジタル信号は、スペクトル データファイル24内に記憶される。

分析結果は、分析結果ファイル25に入力されさらに26では音楽表示を付与する操作が行われ、これは、スピーカ27で音声出力され、又は、ディスプレイ28でグラフ表示される。この表示は第6図に示される。これらの結果は、音楽表示ファイル29に記憶される。

第5B図に示す通り、音楽表示ファイル29内の音楽表示は、最適なスペクトル テンプレートの選択オペレーション30を行う間に、分析結果ファイル25内の分析結果と組合せられる。

この選択オペレーションの結果は、辞書ファイル31に転送される。音楽と辞書内の音楽片との

間の周波数の関連性が3.2で計算され、これをもとに3.3で音楽分布テーブルを作る。

第5図で行われる方法は音楽検出のための特別なプログラムを必要とする。

この方法は、可能な限り正確に、かつ速く容易にセグメンテーションを行う為の、非常に融通性の高いコマンドを持つオーディオグラフ表示フィードバック装置を含んでいる。

第6図のグラフ表示は、時間の関数で波形を示し、同時に周波数、帯域、フォルマント周波数を示している。

プログラムは、音楽及びそのフレームさらにその他音声表示に関する何らかの情報を含む音楽表示ファイルを作り出す。

この後の組のプログラムは、音楽表示ファイルと線形予測符号化法による分析オペレーションの結果を記憶する分析結果ファイル25(第5A図)を、使つて上記のアルゴリズムと関連する音楽の周波数に従いアンプレート解着を構成する。

出力プログラムは結果を分類し、これをヒスト

グラム形式で表示する。

例えば以下で示すような試験用語句のグループは、計算される可能性がたよることがないような、音楽分布を含むように選択される。

#### 試験用語句のグループ

各々が10個の句を含む2つリストは、各々の音楽の発生する数が平均して(平均1.6回)現われるように選択されている。

#### リスト1

- 1 - est-ce que le conducteur arrête l'auto  
(ドライバーは自動車を停止させましたか)
- 2 - c'est toujours comme ça depuis dix ans  
tu sais.  
(10年間もいつもこんなふうだった)
- 3 - ce cheval peut marcher au pas.  
(この馬は常歩ができる。)
- 4 - la bière est moins forte que le rhum.  
(ビールはラムほど強い酒ではない。)
- 5 - ici il fait toujours très froid en hiver.  
(ここは冬は非常に寒い。)
- 6 - J'aime Sylvie quand elle est mignone  
(私は愛くるしいシルビーを愛している。)
- 7 - Diane ne reviendra pas avant lundi.  
(ダイアンは日曜日まで帰ってこない。)
- 8 - aimez-vous le dessin ?  
(絵を画くのが好きですか?)
- 9 - J'ai déjà lu la réponse qu'il m'a  
envoyé par la poste.  
(私は、もう彼が郵便で送ってくれた返事を呼んだ。)

#### 10 - mes gants sont usés.

(私の手袋はくたびれている。)

#### リスト2

- 1 - que désirons-nous pour ce lunch de  
lundi ?  
(月曜日の昼食には何を食べたいか?)
- 2 - faisons vite notre choix, ce sont de  
bonnes choses.  
(早く自分のものを選びこれらはよいものばかりだ。)
- gâtons-nous.  
(気ままにやろう)
- 3 - on guinche ensemble demain, J'aimerais  
du pain, du vin chaud, du boursin,  
du gâteau.  
(明日はいつしよにダンスへ行こう。私はパンやワイン チーズそれにケーキを食べたい。)

4 - Fernand aime voir des feux-follets dans une forêt quand il fait chaud.

(フェルナンドは光が森を暖くつつむとき森の中でかどる光を見るのが大好きだ。)

5 - tu peux causer encore, on ne peut rien faire pour toi, c'est la loi vois-tu.

(あなたは、もつと話すことができるが我々は、あなたに何もすることはできない。)(これが決まりなのです。)

6 - Guy vient de perdre un gant, une guenille, une guêtre, un bonnet, un bandeau, un chapeau, ses lunettes, ses chaussettes, maintenant il grelotte.

(ガイは、手袋も古着もぼうしもヘアバンドメガネソックスをなくしてしまつたので彼は今ふるえている。)

7 - ces bonbons sont bons, on en veut encore tonton Jean.

(これらのキャンディはおいしい。私たちはもつとほしいジャンおじさん。)

8 - que ton chien, ton chat, ton cheval se dépêchent quelles chenilles.

(あなたの犬、猫、馬を急がせなさい。なんてのろまなの！)

9 - bébé joyeux fit joujou puis mangea sa bouillie.

(幸福な赤ちゃんは遊び、それから粉ミルクをのむ。)

10 - suicot fit un oiseau épuisé tombe dans un puit.

(チーチーと鳴きながらつかれた鳥は井戸に落ちた。)

これらの句は、全部で568の音素を含み10msのフレームの63000の音素片からなる。

男性又は女性のアクセントが音調の異なる話者が選択されこれらの語句を登録する。

以下に、第1図の音声認識システム及び第5A図及び第5B図の試験用手段の構成部の実施例を示す。

#### 第1図の音声認識システム

1...A/D変換器+TMS 320 2...約32から64KビットのROMメモリ 3...約16KビットのROMメモリ 4...スタンダードマイクロプロセッサ例えばテキサス・インスツルメンツのTMS 7000 5...RAM又はIPROMメモリ

第5A図及び第5B図の試験用手段

21...A/D変換器 23, 25, 30, 32...ミニコンピュータ例えば適当にプログラムされたテキサス・インスツルメンツのDB 990-12, 24, 25, 28...RAMメモリ 磁気メモリ, 31, 33...磁気メモリ。

#### <効果>

以上の様な構成により要求にあわせた語いの辞書を形成するための登録の作業は簡単になり、かつ音声ユニットへの正確なセグメンテーションが可能になる。

また、本発明の方法によると、音声から音声テンプレートを作り出す作業が容易になる。

従つて音声認識装置の能力は向上し応用可能な範囲が広がる。

#### 4. 簡単な図面の説明

第1図は、不特定話者単語認識システムの主要部分のブロック図である。

第2図は、認識すべき単語と、音素片の形式で表示される辞書の単語とを対応させる方法を示す図である。

第3図は、本発明に従う音声認識の為のベクトル辞書の形成工程を示す図である。

第4図は、テンプレート辞書がそこから形成される領域に音楽表示を付与する工程を示す。

第5A図及び第5B図は本発明に従う不特定話者音声認識の試験手段を示す図である。

第6図は、分析された信号を波形で示したもの及び信号及び信号の音素片表示を構成するパラメータを示す図である。

代理人 浅 村 皓

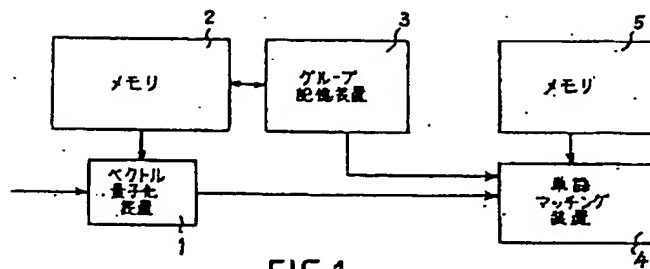


FIG. 1

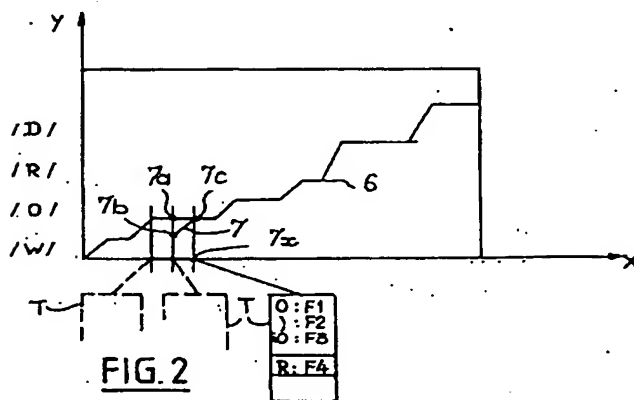


FIG. 2

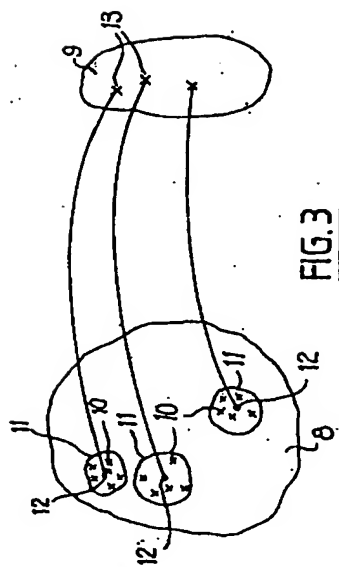


FIG. 3

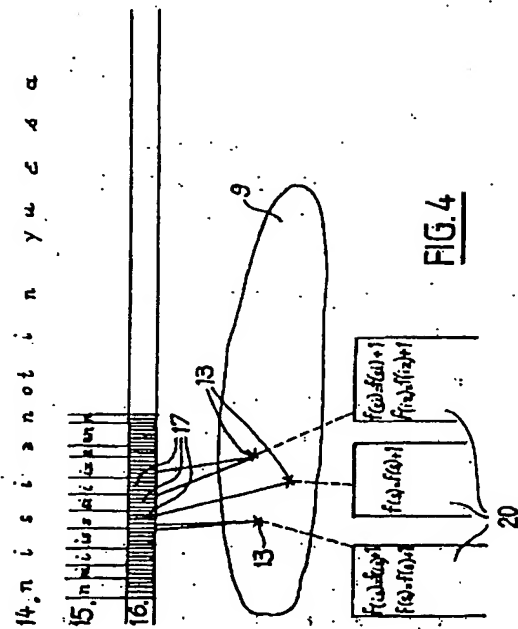


FIG. 4



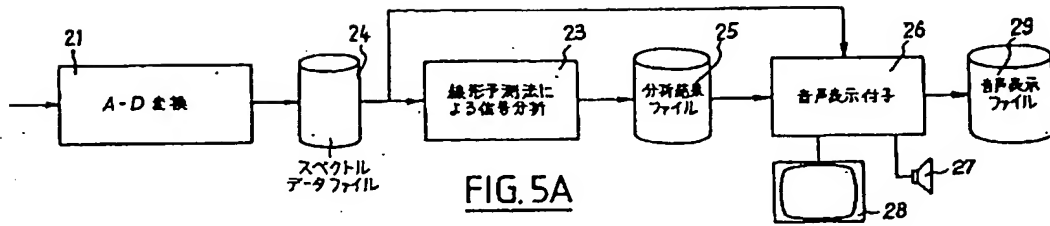


FIG. 5A

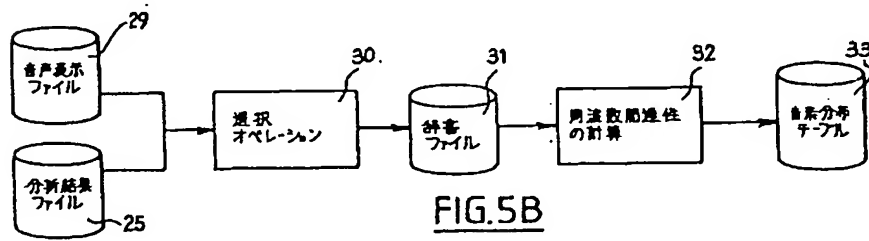


FIG. 5B

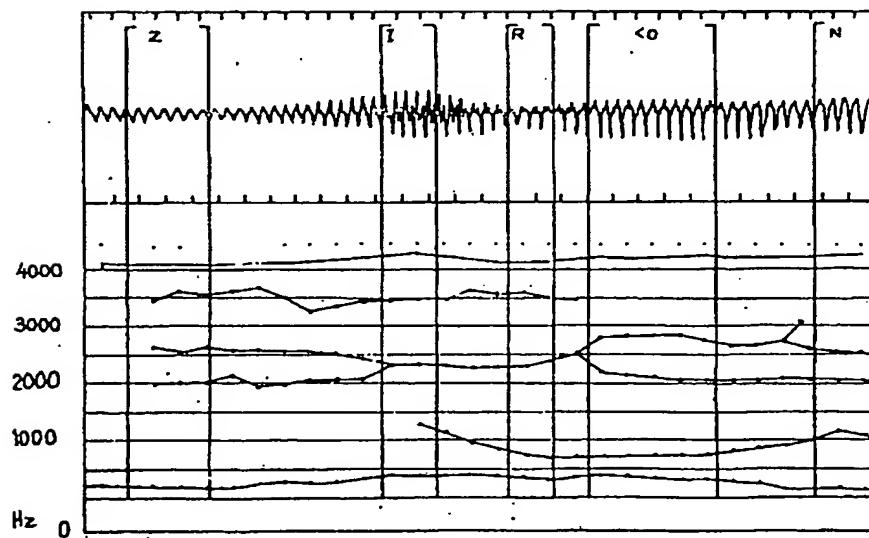


FIG. 6

**THIS PAGE BLANK (USPTO)**

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**